Comments by Bernard Beauzamy on the Financial Times article "How computers killed the expert"

September 3rd, 2007

This article is very interesting for our "Robust Mathematical Modeling" program. It starts with a description : the decisions taken by the Supreme Court, in the United States, were better predicted by mathematical models (based upon simple indicators), than by a panel of legal experts.

We had a similar experience in 2006 in our contracts, with the French Railway, which asked us : will the price of offices continue climbing in the next years ? We built a simple (what we call "robust") model, based upon three indicators : net income per couple of buyers, interest rate, prices of new houses in the United Kingdom. These three indicators, with a time shift of two years, were found highly correlated with the price of offices in France and lead us to the prediction of further climbing of prices.

So, many experts told us : how can you use only 3 simple indicators like that ? You ignore many factors which may influence the final price of offices !

Yes, indeed, we ignored many aspects, which we did not want to take into account (too costly, too long to get information, and so on). We regarded these three indicators as having sufficient predictive value.

The general idea behind such "robust" approaches is that there is no need to go too much into details : of course, many factors may influence the final result, but these factors will already be apparent inside the chosen indicators.

The choice of proper robust indicators is, in itself, also an expertise, and in this respect the title of the article is misleading. This is a whole new area for mathematics, and the rules to be observed are not clear.

A little more technically, I would say that one should not try to find statistical correlation (which is by definition linear) but probabilistic links. Purely statistical tools put into evidence only linear links, and such links are rare. In order to show other dependencies, conditional probabilities are necessary.

Finally, I would certainly fight against using probabilistic tools for situations connected with rare crimes, as the article describes. In France, there was a discussion about the use of such tools for all kinds of juridical situations, and I strongly opposed it. In my opinion, probabilistic tools may be used only if the following two conditions are satisfied (together) : 1) the law of the phenomenon is well understood ; 2) there are enough data.

If the number of data is insufficient, the predictions may lead to absolutely absurd situations. The same if you try to use data collected in one situation in other situations were they do not apply. For instance, to remain in the juridical sector, you may collect data about burglaries, and try to connect them with external indicators (wealth, economy, and so on) ; probably you will get interesting results. But to "extrapolate" to crimes would not be legitimate.